

WHAT IS CLAIMED IS:

1. A method of training a natural language unit comprising:

generating a meaning set from a first corpus using a first natural language unit;

generating a second meaning set from a second corpus using a second natural language unit;

comparing the first meaning set to the second meaning set to generate a score ; and

using the score to determine how to modify the first natural language unit .

2. The method of claim 1 wherein the first corpus comprises a corpus written in a first language and the second corpus comprise the corpus written in a second language.

3. The method of claim 2 wherein the second corpus is aligned with the first corpus. .

4. The method of claim 1 wherein generating a meaning set from the first corpus comprises:

performing a syntactic parse on the first corpus to produce a set of syntactic parses;

09493-044  
T04090-EE864860

performing semantic interpretation of each syntactic parse to produce the meaning set.

5. The method of claim 1 further comprising before using the score:

changing the specification of at least one component in the first natural language unit;

generating a third meaning set from the first corpus using the first natural language unit with the changed specification; and

comparing the third meaning set to the second meaning set to generate a second score.

6. The method of claim 5 wherein using the score to determine whether to modify the natural language unit comprises comparing the score to the second score and modifying the natural language unit based on the difference between the score and the second score to produce a modified natural language unit.

7. The method of claim 5 further comprising after modifying the natural language unit performing steps of:

004090-EE864850

generating a fourth meaning set from the first corpus using the modified natural language unit; comparing the fourth meaning set to the second meaning set to determine a third score; and using the third score to determine whether to further modify the natural language unit.

8. A computer-readable medium having computer-executable instructions for performing steps for training natural language units, the steps comprising:

converting a corpus of sentences into at least two meaning sets using at least two different natural language units; and comparing the meaning sets to evaluate the performance of one or more of the at least two natural language units.

9. The computer-readable medium of claim 8 wherein converting a corpus of sentences comprises converting a corpus comprising sentences from at least two different languages.

10. The computer-readable medium of claim 8 wherein the steps for training further comprise:

FOIA b 7 - EE64360

converting at least part of the corpus of sentences into a modified meaning set using the modified natural language unit; and

11. The computer-readable medium of claim 10 wherein the steps of training further comprise performing the steps of changing the natural language unit, converting at least part of the corpus into a modified meaning set and comparing the modified meaning set to another meaning set for each of a plurality of changes to the natural language unit.

13. The computer-readable medium of claim 12 wherein changing a natural language unit comprises

changing a syntactic parser in the natural language unit.

14. The computer-readable medium of claim 12 wherein changing a natural language unit comprises changing a semantic interpreter in the natural language unit.

15. The computer-readable medium of claim 10 wherein changing at least one natural language unit comprises changing at least two natural language units.

16. A method of training a natural language unit comprising:

generating a first action set from a first corpus using a first natural language unit;

generating a second action set from a second corpus using a second natural language unit;

comparing the first action set to the second action set to generate a score ; and

using the score to determine how to modify the first natural language unit .

17. The method of claim 16 wherein the first corpus comprises a corpus written in a first language

004950"EE64860

and the second corpus comprise the corpus written in  
a second language.

FO4050"EE364860